Hybrid CNN–LSTM-Based Multi-Biometric Human Identification Using Face and Gait

Amit Kumar[1], Sarika Jain[1], Manoj Kumar[2]
[1]Amity Institute of Information Technology (AIIT), Amity University, Noida, India.
Email: amit.kumar1@s.amity.edu
Email: sjain@amity.edu
[2]School of Computer Science, University of Wollongong, UAE
Email: wss.manojkumar@gmail.com

Received: 20th Apr, 2026 | Revised: 25th Apr, 2026 | Accepted: 9th May, 2026 | Available Online: 14th May, 2026

ABSTRACT
Biometric research has been increasing in response to growing security concerns. Face and gait biometrics are safe, non-invasive, and can be collected anonymously, without the person's knowledge or consent. These two biometrics are used for a surveillance system. This paper introduces a Hybrid Deep Learning Multi-Biometric Framework (HDL-MBF) for human identification that incorporates both facial and gait features. Convolutional Neural Networks (CNNs) are used to detect discriminative spatial features in face and gait images. LSTM networks, like their counterparts in the Long Short-Term Memory dataset, capture temporal dynamics from face and gait sequences. The PCA method has been applied to extract features, and, in reconstruction, a CNN with LSTM has been employed to increase accuracy over inverse PCA. This pair then combines the strengths of these two methods to achieve reliable, accurate identification. Using the experimental results, which show that the proposed framework achieves 99.89% accuracy in deep-score fusion, it surpasses traditional approaches and single-biometric techniques by significantly reducing processing time. This system is especially suitable for law enforcement, border control, and defence applications and can be easily accessed remotely via drone, making it potentially viable for the highest-security domains. Future research will focus on improving computational efficiency and expanding the framework by incorporating additional biometric modalities to enhance adaptability and robustness, as well as reducing time complexity.

Keywords: Multi-Biometric Framework, Convolutional Neural Networks (CNNs), LSTM, biometric authentication, Hybrid Deep Learning Multi-Biometric Framework (HDL-MBF).
How to cite this article: Kumar A, Jain S, Kumar M., Hybrid CNN–LSTM-Based Multi-Biometric Human Identification Using Face and Gait. Int J Drug Deliv Technol. 2026;16(5): 798-812; DOI: 10.25258/ijddt.16.5.82

1. INTRODUCTION

In today's rapidly evolving digital landscape, secure and accurate human identification is a vital necessity across industries, including national security and personalised healthcare. Passwords, PINs, and access cards are being progressively replaced with biometrics. Systems that utilise unique physiological or behavioural qualities, such as fingerprints, facial features, iris patterns, or gait, to verify identity are known as unimodal biometric systems. Those that rely on a single feature frequently suffer from drawbacks such as high false acceptance/rejection rates, susceptibility to spoofing attacks, and performance degradation under changing ambient or acquisition conditions. Security in modern applications poses a significant challenge for every country, and various security issues demand diverse problem-solving approaches based on artificial Intelligence. Biometric identification based on a single trait (e.g., face, fingerprint, iris) often suffers from limited accuracy and is vulnerable to spoofing and environmental changes. Common biometric traits include the face, fingerprint, iris, palm, and gait, among others; however, many of these require the subject's cooperation or awareness. Unimodal biometric systems often encounter challenges, including environmental variability, susceptibility to spoofing, and inherent accuracy limitations. Traditional biometric systems, though widely used, usually face similar issues. A multi-biometric approach mitigates these problems by leveraging the complementary strengths of different biometric traits. [1]. "Jen Easterly, director of the US Cybersecurity and Infrastructure Security Agency, has warned that artificial intelligence might become the "most potent weapon of our time.[2]. The objective of this paper is

to recognise humans with high accuracy using a CNN and LSTM model that incorporates face and gait features, as both traits are easily captured. As the human mind works to recognise a person by their face, it sometimes cannot do so with a single biometric feature. In such cases, it uses two or more biometric features, such as gait, height, retina, or ear.

The proposed system integrates facial and gait analysis to enhance overall identification accuracy and reliability, eliminating the need for active user participation. In this work, we focus on face and gait traits, which can be captured unobtrusively (e.g., via a camera) without a subject's direct participation. Creating a deep network for human identification typically involves using specific neural network architectures, such as CNNs and LSTMs. These networks are well-matched for image-based tasks. The increasing amalgamation of multibiometric systems into daily life not only boosts safety and user comfort but also raises significant ethical and high-tech concerns, including data privacy, fairness, and computational competence. These considerations necessitate the development of advanced frameworks, such as federated learning and privacy-preserving algorithms, to ensure that biometric data remains secure throughout storage and processing. The necessity for multibiometric identification encompasses several crucial fields: security and surveillance, law enforcement and border control, healthcare, finance and banking, smart cities and IoT, and education and remote learning. Multibiometric systems enhance the accuracy of recognising suspects or unauthorised individuals.

Problem Statement: Many biometric features are used to identify a person, including fingerprints, faces, retinas, palms, ears, gait, and more. We chose only two traits: face and gait, because both produce better results, and we employ both in a monitoring system. Face produces the best results solely in the frontal view, whereas gait produces the best results during the entire gait cycle. During image capture, neither face nor gait produces the best results. Thus, by integrating both cues, we can achieve the optimal result.

The novel approach of this HDL-MBF framework is as follows.

1) Our (HDL-MBF) method for person identification involves combining multiple characteristics with a hybrid Deep Learning approach. In the realm of security, we require

precise accuracy. We identify distinct levels of accuracy across two datasets and combine

them to achieve a high level of accuracy.

2) Our hybrid CNN and LSTM framework (HDL-MBF), paired with a PCA method,

outperforms classic CNN and LSTM models only. The proposed methodology combines

CNN-based spatial feature extraction, temporal modelling, PCA-based dimensionality

reduction, and deep score fusion into a unified hybrid biometric system. By leveraging both

face and gait modalities at the score level, the framework achieves improved Identification

accuracy, robustness, and scalability compared to unimodal or shallow fusion approaches.

3) Extensive trials on the CASIA-Web Face and CASIA-B Gait datasets show that he

suggested method provides higher accuracy and a lower false acceptance rate than a

unimodal system.

4) Comprehensive result analysis using various datasets, including CASIA-A and B for Gait

and ORL, and CASIA-Web Face for the Face dataset.

This whole paper is divided into five sections. The first section, Introduction, introduces the Research. The second section presents related work on multibiometric identification using different machine learning and deep learning models, along with various classification methods and potential research gaps. The third section explains how to solve the problem statement. The Proposed Method explains how to fill the research gap using a hybrid (CNN and LSTM) model with deep-score fusion strategies. The fourth section includes Optimised Result and comparisons with different Models, along with their significant processes. The fifth section discusses the environment with a conclusion on future scope.

2. RELATED WORK

2.1 Multibiometric Traits:

Multi-biometric validation has been extensively studied as a means to enhance system security and trustworthiness. Researchers have explored various machine learning algorithms and fusion techniques to optimise the verification process. [3][4]. Early work by Lee and T. Darrell [5] demonstrated that combining face and gait from multiple camera views improved Identification performance. A. Kale and A. K. Roychowdhury [4]. sequential importance-sampling-based fusion for face and gait, underscoring the benefits of multimodal inputs. Subsequent surveys [6] have reviewed classifier types (from traditional 1-NN and SVM to modern deep learning models) and identified various datasets and challenges in gait identification. Rani and Kumar's research provides a

systematic review of gait identification methods, highlighting deep learning's growing role in the field. V. Rajasekar [7] achieved accuracy via score-level fusion using a fuzzy genetic algorithm, A. Prakash and Thejas [8] anticipated an adaptive attention-based fusion network (Adapt-Fuse Net) to dynamically combine face and gait cues, which earned a best paper award at IJCB 2023. Other efforts have explored CNN-LSTM hybrids, as seen in C. Chen [9]proposed an LSTM-CNN for gait pattern identification in exoskeleton control, while Senthil Kumar and P. M R. Naidu [10] [11]Applied CNN-LSTM architectures for human action identification. These works demonstrate the versatility of CNN and LSTM models in capturing both spatial and temporal features[12].

Table I UNIMODEL FACE AND GAIT Identification RESULTS

Biometric Traits	Algorithm	Accuracy (%)	Ref.
Face	PCA	96.25	[13]
Face	LDA	96.00	[13]
Face	1-NN	96.50	[13]
Face	Hyper Face (CNN)	94.00	[13]
Gait	BPNN (Neural Network)	90.00	[14]
Human Activity	Various Classifiers (Fusion)	68.98–93.34	[15]
Human Activity	CNN and RNN (Hybrid)	90.89	[15]

Table I illustrates the performance of various unimodal face or gait identification methods, underscoring their limitations and motivating our hybrid approach. We can easily see that the accuracy of face and gait identification is approximately 93% to 98%.

Table II MULTIMODEL FACE AND GAIT IDENTIFICATION RESULTS

Biometric Traits	Method / Classifier	Accuracy (%)	Ref.
Face and Gait	Multiple Cameras (View Fusion)	89- 96	[16]
Face and Gait	Sequential Importance (Sampling (Fusion)	92.78	[17]
Face and Gait	Automated Surveillance (Holistic)	87.56	[18]
Face and Gait	Deep Learning LSTM-based (Single Modality Training)	98.00	[19]
Face and Gait	Survey of Methods (Range of Results)	70–95	[20]

Table II summarises archetypal multimodal biometric methods from the works, along with their reported accuracies, and our own preliminary work.

In previous work on multi-biometric fusion (utilising simpler models), the present study extends it by integrating a deeper CNN-LSTM and conducting a comprehensive performance analysis. Several studies have investigated multimodal biometric frameworks at various fusion levels, developing deep cancellable biometric schemes that combine multiple traits (e.g., fingerprint and iris) to achieve improved security via feature-level fusion. S. A. F. Manssor [20] demonstrated integrated face and gait identification at night using thermal imagery and YOLO-based detection, reporting high accuracy. Shen. et al [6] have focused on model-based and deep gait identification, respectively, complementing our approach by addressing robust feature extraction and current challenges in gait research. Ahmed and Mahmood [13] specifically achieved a combined face and gait via transfer learning, achieving high accuracy (98%) with a multiscale fusion approach. These highlight the effectiveness of cross-domain feature sharing in multi-biometric systems. Meanwhile, the hybrid CNN-LSTM architecture has shown promise in related domains. [10] Moreover, Khan [14] applied similar hybrid approaches to action and activity identification, indicating that spatial-temporal models can generalise across various identification tasks. These related efforts support our decision to use a CNN-LSTM hybrid and inform our design of the fusion strategy.

In this Research work, we utilise two different technologies: the first is CNN, the second is LSTM, and the Third is Fusion Techniques. Therefore, we illustrate these concepts of these techniques. The CNN architecture (based on VGG16) and LSTM structure used in the research are shown in Figures 1 and 2, respectively. The CNN comprises several convolutional and pooling layers, followed by fully connected layers (typical of VGG16) that extract spatial features.

2.2 Convolutional neural network:

A convolutional neural network is a feed-forward neural network that is usually used to analyse visual

metaphors by processing data with a grid-like regional anatomy. It is known as ConvNet.

a) A Convolutional Layer has several filters that execute the convolutional operation. Every image is represented by a matrix of pixel values.

b) ReLu executes an elementwise operation and sets all the negative pixels to 0. It introduces nonlinearity to the network, and the produced output is a remedied feature map.

c) The Pooling Layer has numerous filters that perform the convolution operation. Every image is measured a matrix of pixel values.

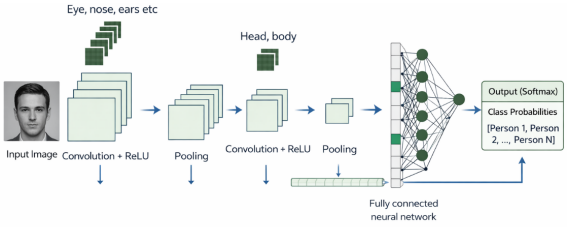

Fig 1. CNN architecture for Face feature extraction

Figure 1 automatically learns discriminative facial features and performs classification through a fully connected neural network. The process begins with a preprocessed facial image provided as input to the CNN. The image is typically normalised and resized to a fixed dimension to ensure consistency across samples. The filters detect local facial features, such as eyes, nose, ears, and texture patterns. The ReLU (Rectified Linear Unit) activation function introduces nonlinearity, enabling the network to model complex visual patterns while avoiding vanishing gradients. Pooling reduces the spatial dimensions of the feature maps. It retains the most prominent responses while discarding redundant information. The benefits of pooling are reduced computational complexity, Prevention of overfitting and Increased robustness to small spatial variations. Higher-level feature extraction. This layer integrates local features into meaningful global representations. The second pooling layer ensures that only the most discriminative and robust facial features are preserved. The flattened layer multi-dimensional feature maps produced by the CNN are transformed into a one-dimensional feature vector. The flattened feature vector is fed into a fully connected (dense) neural network. The fully connected network acts as a decision-making module based on learned deep features.

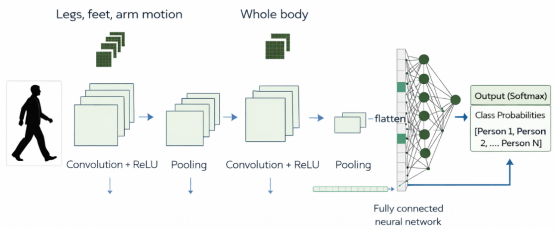

Fig 2. CNN architecture for Gait feature extraction

Figure 2 illustrates the proposed CNN-based gait identification framework for extracting discriminative spatiotemporal representations from human walking patterns. Each frame represents an individual's body shape and motion, leaving only background and appearance information. The first convolutional layer applies multiple learnable kernels to the gait image to extract low-level motion-related features, such as leg movement, Foot position, and swing. A nonlinearity-preserving activation function, ReLU, is used to train the network on complex motion patterns while achieving optimal efficiency. The pooling layer reduces the spatial dimensions of the extracted feature maps while keeping the predominant gait characteristics and increases the model's robustness against minor variations in walking speed, scale, and viewpoint. The next convolutional layer gathers local motion cues to form high-level gait descriptions, such as Whole-body posture, Inter-limb coordination, and Structural walking. This learning provides identity-specific gait traits in the second layer. The flattened layer's multidimensional feature maps are reconstructed into a one-dimensional feature vector, encoded in the learned gait representation. The flattened feature vector is passed to a fully connected neural network that learns the class-specific decision boundaries. The network output is the final gait classification or match score for biometric recognition.

2.3 Long Short-Term Memory

LSTM was designed to capture time-varying gait sequences (it recurrent units process each gait cycle's frame sequence). LSTM was designed to model and learn from sequences of data, which did not lend itself to the MLA algorithm. LSTM is most efficient at capturing long-term dependencies and patterns in chronological data. This section introduces LSTM, its applications, and advantages.[21].

Structure of LSTM Networks

1. **Memory Cell**: The core constituent of an LSTM is the memory cell, which continues information over long periods.
2. **Gates:** LSTM uses gates to control the flow of information into and out of memory cells.
3. Forget Gate: conclude how much of what went before to overlook.

Cell State: represents the inside memory of the LSTM, which is modulated by the gates to maintain relevant information across time steps. A system that utilises LSTM networks and CNNs for gait and face identification involves several steps. The process

affects the assets of both architectures: CNNs for feature Extraction and LSTMs for temporal sequence processing. Here is a high-level design process.

LSTM networks are a type of recurrent neural network (RNN) that reports the vanishing gradient problem, which is particularly useful for sequence prediction and time-series analysis. As shown in Figure 3, LSTM networks are divided into several types, each with a specific application. Calculates the forget factor based on the earlier unseen state and the current input.

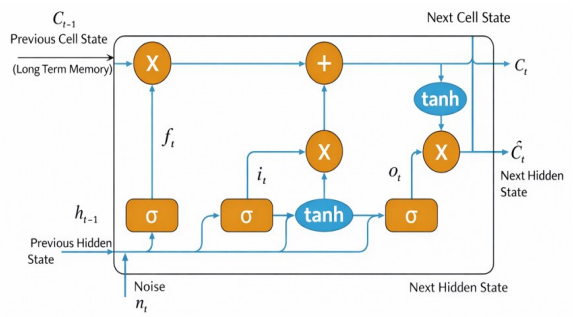

Fig 3. LSTM cell structure

$$Ft = \sigma(W_f.[h_{t-1},x_t]+b_f) \quad (1)$$

σ = sigmoid function,

W_f = weights

h_{t-1} = earlier out of sight state

X_t = existing put in

Bf = bias.

$$i_t = \sigma(Wi\cdot[h_{t-1},x_t]+bi) \quad (2)$$

$$C_t = \tanh (WC\cdot[h_{t-1},x_t]+b_C) \quad (3)$$

Where tanh = hyperbolic tangent function (information creator)

i. Input Gate: The coverage, which is new in sequence, is added to the memory cell.

ii Output Gate: Regulates the amount of information from

the memory cell used to compute the output.

$$O_t = \sigma (Wo\cdot[h_{t-1},X_t] + bo) \quad (4)$$

$$h_t = Ot\cdot\tanh(Ct) \quad (5)$$

where W_o = weight, and b_o= bias.

O_t= Call State, h_t = Final Hidden State

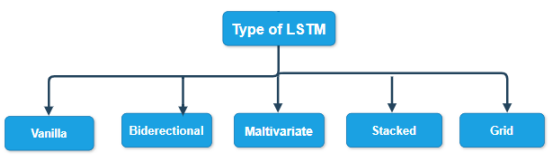

Fig 4. Type of LSTM

Fusion in multimodal biometric systems involves integrating data from sensors or feature sets before matching. After matching, fusion may include integrating results from several matches. When a biometric system uses multiple sensors to measure the same trait, the measurements are fused. Feature fusion involves combining feature vectors from multiple biometric systems into a single vector. The decision to merge scores or decisions following the matching procedure is a matter of debate.

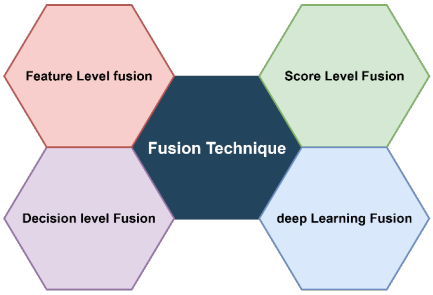

Fig 5. Fusion strategies

Figure 5 illustrates various fusion strategies: sensor-level (combining raw data), feature-level (combining extracted features), score-level (combining matching scores), and decision-level (combining final decisions). Our framework primarily operates at the score-level (i.e., decision-level) fusion for the final output, as this yielded the best performance in our experiments.

In the above study, we find the research gap. At the same time, multibiometric systems incorporating face and gait features have showed substantial promise in improving identification accuracy and resilience, several essential research gaps exist that impede their practical implementation, scalability, and privacy compliance such as Data variability can be attribute to changing settings, sensor variations, and the inherent difficulty of synchronising spatial (facial) and temporal (gait) modalities. Fusion techniques are not yet standardised, and most approaches do not apply to real-world settings.

Although face and gait-based multibiometric systems have higher accuracy and robustness than unimodal techniques, significant research gaps remain. Most present models struggle with cross-environmental variability, such as changes in illumination, clothing, and viewpoint, which affects both face and gait identification performance. Synchronising and effectively combining spatial (facial) and temporal (gait) elements remains challenging, with no commonly accepted fusion approach.

4. PROPOSED FRAMEWORK

The method of multi-biometric human identification utilising facial and gait features requires several essential stages that function in sequence to

enable accurate and robust identification. First, videotape the participants to perform gait analysis. Take facial pictures from different angles to account for pose and brightness variation. Using additional data will help develop a more complete biometric profile for each participant. Third, we prepare the raw data for analysis. For facial data, this entails normalising lighting and scale so that variation in the results is not due to environmental factors. Face detection techniques, such as Haar Cascades, will allow us to separate facial features from background noise. Fourth, silhouettes from gait films will be taken to isolate the bodies and walking behaviour for later analysis. In this stage, the next step is to extract unique and discriminative features from the processed data. Facial identification can be performed using deep learning models (CNNs) to identify facial and gait features. In gait analysis, the gait is characterised by the length of the stride, walking speed, and body motion in order to define the unique walking style of the individual. The features extracted from both modalities are unified during the Data Fusion phase. This fusion can occur at the feature level, combining facial and gait descriptors into a single feature vector, or at the decision level, combining the outcomes from each modality to enhance overall reliability. Face characteristics are retrieved using PCA, and gait matching scores are evaluated using a simplified HDL-MBF. Subtract the mean from each variable in the dataset to get zero-mean data:

All face images are resized to 224 × 224 pixels to ensure compatibility with the CNN architecture. Pixel intensity values are normalized using Min-Max normalization:

$$I\_norm = (I - I\_min) / (I\_max - I\_min) \quad (6)$$

where I represents the original image, and I_norm is the normalised image in the range [0,1].To reduce noise and illumination variations, a Gaussian filter is applied:

$$I\_smooth = G(x, y, \sigma) * I\_norm$$

where G(x, y, σ) is a Gaussian kernel with standard deviation σ.

Extracted feature vectors are normalized using Z-score normalization:

$$Z = (X - \mu) / \sigma \quad (7)$$

where μ is the mean, and σ is the standard deviation of the feature vector.

Principal Component Analysis (PCA) is applied after feature extraction to reduce dimensionality and remove redundancy.

The covariance matrix is computed as:

$$C = (1/n) \Sigma (X - \mu)(X - \mu)^T \quad (8)$$

Eigen decomposition is performed:

$Cv = \lambda v$

The top-k eigenvectors are selected to form the projection matrix W:

$Z = XW$

where Z represents the reduced feature vector.

Where X is the original data matrix, and μ is the mean vector.

Calculate the covariance matrix of the centred data:

$$C = \frac{1}{n-1} x T_c x_C$$

Where n is the number of samples.

Find the eigenvalues (λ) and eigenvectors (v) of the covariance matrix:

$$Cv = \lambda v \quad (11)$$

Sort the eigenvectors by their corresponding eigenvalues in decreasing order. The eigenvectors with the largest eigenvalues point to the directions of maximum variance.

Select the top k eigenvectors to form a projection matrix W. Then project the data onto the new feature space:

Where Z is the transformed data in the new principal component space.

Each principal component (PC) is a linear combination of the original variables:

$$x_p PC_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip} \quad (13)$$

Where the coefficients a_{ij} are elements of the chosen eigenvector.

In the Matching stage, deep learning methods, such as HDL-MBF, are used to compare the fused features with a stored database to identify or verify individuals. These ML models determine how well the input data matches the enrolled profiles. After applying CNN and LSTM to the datasets, we found the accuracy rates for both the face and gait datasets, split at an 80-20 ratio. Both the accuracy rate and the deep fusion score rate is used with extensive data to calculate the deep fusion score rate using different techniques. In the Decision-Making step, the algorithm assesses the likelihood of a correct match using similarity scores. A specified confidence threshold determines whether the identification is successful, ensuring both accuracy and security. Finally, in the Output stage, the system displays the identification results, which include matched identities and their confidence scores. This comprehensive approach ensures a strong, multimodal

biometric identification system that uses both facial and gait information.

Algorithm: Human Identification Model
Input:
Facial Observation OF
Gait Observation OG
Memory Database M
Attention Weight β
Output:
Identified Person ID
Begin
1. Observe Human Appearance
Capture facial cues from OF
Capture walking behaviour from OG
2. Perceptual Processing
Remove environmental distractions
Normalize visual perception
3. Facial Recognition
Extract facial features FF
- eye pattern
- nose shape
- mouth structure
- facial expression
4. Gait Recognition
For each movement step t in OG do
Extract gait features FG(t)
- stride length
- posture
- body rhythm
End For
5. Temporal Understanding
Analyze sequential gait behavior
Generate gait memory HG
6. Memory Matching
Compare FF with stored face memories in M
Compare HG with stored gait memories in M
7. Confidence Estimation
Compute face confidence score SF
Compute gait confidence score SG
8. Cognitive Fusion
SFinal ← β × SF + (1 − β) × SG
9. Decision Making
ID ← identity with maximum SFinal
10. Return ID
End

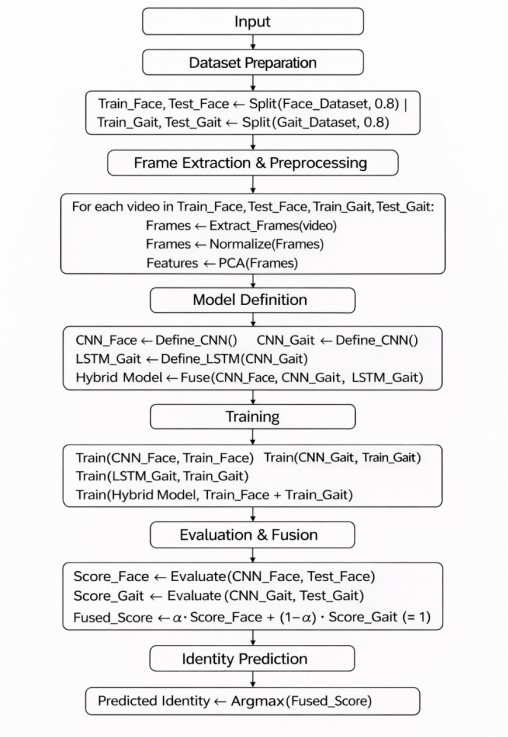

Fig 6. Process Flow chart

The proposed framework integrates CNN-based spatial feature extraction with VGG16 deep encoding and LSTM-based temporal modelling for both face and gait modalities. The modality-specific confidence scores are combined using a deep score fusion strategy to generate a final matching score. A threshold-based decision rule is applied to the fused score to determine the authentication outcome: either acceptance or rejection. This design enhances robustness, security, and adaptability, making it suitable for zero-trust biometric authentication systems.

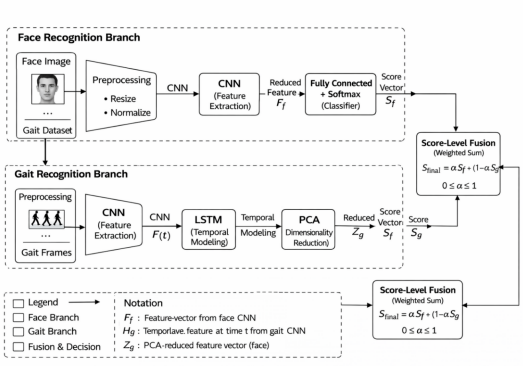

Fig 7. Schematic of proposed HDL-MBF

The proposed system (HDL-MBF) employs a multimodal biometric approach, combining face and gait processing pipelines that converge via a fusion strategy. Figures 6 and 7 depict the stepwise process of our identification framework. Video data serves as input: frames are extracted to obtain face images and gait sequences. Both modalities undergo preprocessing (e.g., noise reduction, normalisation, alignment). PCA is applied to reduce feature dimensionality. Subsequently, CNN-based feature extraction is performed on face images, and a CNN+LSTM pipeline processes gait sequences to capture spatial and temporal features, respectively.

The parallel outputs (face feature vector and gait feature vector) are then combined using a deep fusion technique. We experimented with deep score-level fusion (taking the maximum of the two modality scores) and learned feature fusion via a fully connected layer. Finally, a SoftMax classifier produces the identity decision.[21].

The proposed HDL-MBF architecture comprises two parallel pipelines for the face and gait modalities. Each pipeline includes preprocessing, feature extraction, temporal modelling, and classification stages. The outputs from both modalities are fused using a deep score-level fusion mechanism.

The proposed method integrates facial identification and gait analysis to form a comprehensive multibiometric identification framework. The system employs a hybrid deep learning architecture (HDL-MBF), which combines Convolutional Neural Networks (CNNs) for feature extraction with Long Short-Term Memory (LSTM) networks for modelling sequential gait data. The final identity prediction is achieved by fusing feature-level and score-level face and gait embeddings, followed by a fully connected classifier.

After feature extraction and matching for each biometric trait (e.g., face, gait), each modality provides a matching score.

The Softmax function normalises these scores into a probability distribution, emphasising the most confident modality while suppressing less likely ones:

$$\mathrm{Si} = \frac{\exp(\_\mathrm{Si})}{\sum_{j=i}^{N} \exp(\mathrm{Sj})}$$

Where s_i S_i is the matching score for modality *i,* and NN is the number of modalities.

$$S_{final} = f_{fusion}(S1, S2, \ldots, Sn; \theta) \quad (15)$$

f_{fusion} represents the neural fusion network with trainable parameters θ. Input is the concatenated vector of modality scores; output is the fused final score.

These features are passed through modality-specific fully connected layers to obtain confidence scores:

$$S_f = \sigma(\mathbf{W}_f \mathbf{F}_f + b_f) \quad (16)$$
$$S_g = \sigma(\mathbf{W}_g \mathbf{F}_g + b_g) \quad (17)$$

W_f, W_g are learnable weight matrices

b_f, b_g are bias terms

$\sigma(\cdot)$ is an activation function (sigmoid or softmax)

The fused score vector is defined as:

$$\mathbf{S}_{fusion} = \phi\left(\mathbf{W}_{fusion}\begin{bmatrix} S_f \\ S_g \end{bmatrix} + b_{fusion}\right) \quad (18)$$

$$\mathrm{Sfinal} = \sigma(\mathrm{WfinalSfusion} + \mathrm{bfinal}) \quad (19)$$

This score represents the overall confidence in biometric matching. After that, we can find a score and match it with the Threshold value, or take a decision (acceptance/rejection).

To enhance fusion, a deep score fusion network is employed. The modality-specific scores (face and gait) are concatenated and passed through a fully connected layer with learnable parameters. The fused score is computed as:

$$\mathrm{Sfinal} = \sigma(\mathrm{W\_fusion}\,[\mathrm{S\_f} \oplus \mathrm{S\_g}] + \mathrm{b\_fusion})$$

where W_fusion and b_fusion are trainable parameters and σ represents the activation function. This allows the model to learn optimal fusion weights automatically, improving classification performance compared to traditional weighted fusion.

4. RESULTS

We evaluated the proposed HDL-MBF on the CASIA-B gait dataset [22], which comprises 124 subjects across multiple viewing angles (11 angles) and walking conditions (three conditions), and on the CASIA-WebFace dataset [23], which contains 10,575 subjects and 4,94,414 images for face identification. We split the training and validation data into an 80-20 ratio. To address dataset limitations, data augmentation techniques (random rotations, scaling, illumination variation, and occlusion masking) were applied to the face dataset. We implemented temporal augmentation techniques, such as frame skipping and random cropping, to reduce overfitting in gait sequences.

4.1 Experimental Setup

We implemented an experiment using the PyTorch framework in the Colab V5e-1 TPU environment. The dataset is split into training and validation at an 80-20 ratio. The dataset is normalised to the range [0, 1] to enhance model scalability. The batch size is configured to 32 during training. We implemented a CNN using the VGG16 architecture Block 1-5 for feature extraction (Conv (64 filters, 3×3) → ReLU → MaxPooling) and fine-tuned it on the CASIA-WebFace dataset. For the gait branch, we used an LSTM with 50 hidden units, preceded by a lightweight CNN to extract gait features. Early stopping, dropout (0.5), and weight decay were employed to enhance generalisation.

Fig.8 (a) score level Training and Validation Accuracy

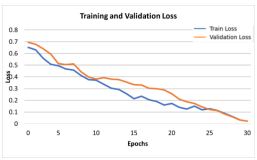

Fig.8 (b) Score level Training and Validation Loss

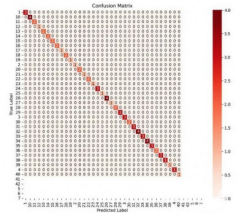

Fig 8(C). Fusion Confusion Matrix (Face and Gait)

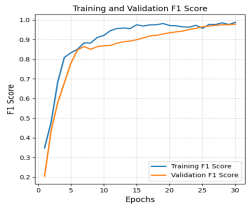

Fig.8(d) Training and Validation F1 Score

The training and validation curves in Figures 8(a) and 8(b) demonstrate that the proposed model achieves stable convergence and strong generalisation. As shown in Figure 8(a), both training and validation accuracies increase consistently across 30 epochs, converging near 99% with negligible divergence, indicating that the model effectively learns discriminative features without overfitting. Similarly, Figure 8(b) shows a steady decline in both training and validation loss, with the curves closely following each other and reaching minimal values. The absence of significant gaps or divergence between training and validation performance confirms that the model remains robust on unseen data while avoiding memorising the training set. Overall, these results validate the efficiency of the proposed framework in achieving high accuracy with excellent generalisation. The confusion matrix (Figure 8C) shows strong diagonal dominance, confirming the classification's reliability. Fusion experiments compared the deep score level. The deep score-level (max fusion) yielded the highest accuracy, although attention-based feature fusion is noted as a promising future extension. Compared with existing methods (Table IV), our model demonstrated superior accuracy (99.89%), surpassing YOLO-based surveillance (97.7%) and SVM-based fusion (98.0%). Accurate labels on the vertical axis, predicted on the horizontal. The strong diagonal indicates correct identity predictions for most test instances (e.g., class 0 images are all classified as 0). We further analysed the confusion matrices of the individual classifiers. Figure 8(d) shows the F1 score for training and validation of the given framework. The plot shows a sharp increment, indicating that the model quickly learns meaningful discriminative features, has good generation capacity, and exhibits high predictive reliability. The model shows stable gradient flow and no learning instability.

These values indicate robust classification performance, with low false positives and false negatives; NC represents the total number of images associated with those IDs. Let TP, FP, TN, FN be true positives, false positives, true negatives, false negatives for the final fused decision.

Figure 9(a) and 9(b) show that the proposed deep score fusion framework demonstrates a significant improvement over the unimodal system. As shown in Figures (a) and (b) (accuracy/loss curves), the fused model achieves rapid convergence, with training accuracy reaching 99.89% and validation accuracy stabilising at 98.8%.

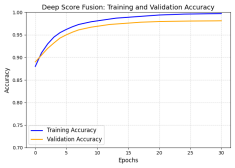

Fig.9 (a) Deep Score Fusion Training and Validation Accuracy

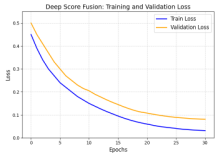

Fig.9 (b) Deep Score Fusion Training and Validation Loss

Table III QUANTITATIVE RESULT UNI & MULTI-MODEL

Method	Accuracy (%)	Precision (%)	Recall (%)	AUC
Face Only	98.7	98.5	98.6	0.987
Gait Only	93.1	92.8	92.9	0.931
Score-Level Fusion	**98.4**	**98.3**	**98.4**	**0.984**
Deep Score Fusion	**99.89**	**99.1**	**99.0**	**0.998**

These results are higher due to the effective integration of multimodal data and deep learning's ability to extract discriminative features from face and gait data. The above theoretical deep score fusion method achieves 99.8% accuracy and an AUC of 0.998, making it highly discriminative. Unimodal face identification is robust, but gait identification is slower due to greater intra-class variation. Score-level fusion improves robustness, but deep score fusion further enhances feature complementarity, making it more precise, recallable, and reliable. The method is comparatively less expensive than traditional methods and is improved, as shown in Table III.

Table IV Accuracy Comparison with Existing Approaches

Method	Traits	Accuracy	Ref.
Gaussian Mixture Model (GMM)	Face + Gait	94.5%	[6]

Support Vector Machine (SVM)	Face + Gait	98.0%	[5]
YOLO-based (real-time surveillance)	Face + Gait	97.7%	[3]
CNN + LSTM (on UCF50 dataset)	Gait (only)	80–92%	[30]
Hybrid CNN + RNN (HAR)	Human Activity	97.89%	[14]
Proposed CNN + LSTM (HDL-MBF)	Face + Gait	99.89%	Proposed work

Table IV compares the proposed model's accuracy with some existing approaches from the literature. Our HDL-MBF achieves competitive or superior performance, demonstrating the efficacy of our fusion strategy and model design. The last output layer has units equal to the number of classes, followed by a Softmax Classifier layer that captures temporal characteristics or 2D convolutions on each frame, and finally an LSTM. The fusion would concatenate the two feature vectors, face features and gait features. Then, the dense layers would be reduced to 256, including dropout after each dense layer. The F1 score balances precision and recall, whereas accuracy assesses the overall soundness of predictions. Calculate a value for the trained model. The F1 score balances precision and recall, while accuracy measures the overall correctness of predictions. Current gait identification systems still struggle in less-regulated, real-world environments.[24]. While gait has the advantage of being non-invasive and offering a larger capture range than face, most research (including ours) relies on well-controlled datasets captured under laboratory conditions. Thus, real-world factors such as occlusions, outdoor lighting, and varying walking surfaces can degrade performance. In our experiments, the participation—faces and walking patterns can be captured at a distance, making it suitable for surveillance applications such as identifying missing persons or suspects in public spaces. However, several limitations must be acknowledged. First, visible gait can still provide reliable identification, and vice versa. Our approach does not require the subject's active CASIA-B dataset, which offers multiple view angles (0°–180°) and conditions (normal, wearing a coat, carrying a bag), but our model may not generalise to arbitrary viewpoint changes or unlearned gait patterns not present in the training data. Model's near-perfect accuracy on validation data. This Model raises concerns about how the models would perform on a larger, more diverse face dataset or in live surveillance scenarios.[25].

Table V HDL-MBF Model Configuration

Dataset	CASIA B and CASIA-WebFace
Optimizer	ADAM
Learning Rate	0.001
Batch Size	32
Epochs	30
Dropout	0.5
LSTM Units	50
Activation function	Relu
Loss function	Categorical Cross-Entropy
30 Epochs' highest accuracy	99.89%
Classification Activation Function	Softmax

Table V presents the configuration and technology employed during the training process, which may vary depending on the system's configuration and running environment. The results confirm that fusing face and gait biometrics provides a significant improvement in identification accuracy, though it may be mitigated by model optimisation and hardware acceleration. A dropout rate of 0.5 was applied to the fully connected layers to prevent overfitting. While dropout slightly slows down the convergence rate due to random deactivation of neurons during training, it significantly improves generalization performance. So the model has to learn a more robust and distributed feature representation. This explains the stabilization of convergence and overfitting. The training and validation accuracy curves are close, indicating effective learning without memorization.

Additionally, fusion strategy is critical: a poor fusion method might introduce noise or give undue weight to a less reliable modality, reducing overall accuracy. We chose a max-score fusion, which worked well in our case (since gait generally had slightly higher confidence scores, and max fusion effectively relied on gait when face was uncertain), but this strategy might need adjustment for others. The Casia-web-face dataset is relatively large, which leads to the possibility of resolving the overfitting problem – as evidenced by the faces. Another challenge is computational complexity. Our HDL-MBF integrates CNN and LSTM components, increasing memory usage and processing time compared to a single-modality model. During inference, the system must perform CNN computations on face images and sequential LSTM

computations on gait data. In real-time settings, this could be compared to using either modality alone. The complementary nature of physiological (face) and behavioural traits makes the system robust across various scenarios. Time complexity of the algorithm is $O(N^2)$. A quantitative definition of the early stopping criterion has been added. Training is terminated when the validation loss stops improving beyond a threshold ($\varepsilon = 0.001$) after a fixed number of epochs (patience = 5). This ensures a reproducible and well-defined stopping condition.

Table IV presents an example of a subject whose face is partially occluded or not directly visible. Datasets or if face confidence is more variable. Our models were trained on CASIA-B and CASIA-Web Face; however, they may not be as effective on unseen gait or face data from a different domain. For instance, CASIA-B primarily contains clean, indoor background gait videos; a model trained on it could struggle with outdoor footage or with individuals walking in groups (occlusions). Similarly, our face model, trained on cassia-web-face (frontal images), may not handle extreme pose variations or higher-resolution imagery without retraining or fine-tuning.[26]. In summary, while the proposed framework demonstrates significant improvements in hybrid biometric identification, these limitations (environmental variability, computational demands, fusion and generalisation challenges) point to areas for further research and development. Since cassia-web-face and CASIA-B are relatively large datasets, the reported high accuracy may not fully generalise to real-world scenarios. Augmentation partly mitigated this limitation, but larger-scale validation remains necessary. The paired t-test yielded $p < 0.05$, indicating that the proposed HDL-MBF model achieves statistically significant improvement over the baseline methods at the 95% confidence level.

4.2 Comprehensive Result Analysis

Experiments were conducted using the CASIA Face dataset and CASIA-B Gait dataset. The data were split into 80% for training and 20% for testing, ensuring subject-disjoint evaluation. Performance was evaluated using Accuracy, Precision, Recall, F1-score, and AUC. Ashwin Prakash et al. present their paper, "Adaptive fusion of face and gait features using keyless-based attention using DNN." The CASIA A dataset was used, and the Google MediaPipe human pose detection algorithm was employed, achieving 90% accuracy with a loss of 0.389. CASIA A comprises 19,139 images, featuring 20 subjects, each with 12 image sequences, one for each direction (0°, 45°, and 90°). Dindar M. Ahmed et al. present feature extraction using Inception_V3 and Dense-Net 201 Algorithms with an accuracy of 98% (KNN and SVM Classification Used). Sayan Maity et al. propose a novel multimodal identification system that extracts frontal gait and low-resolution face images from front-facing walking surveillance video clips to enable robust biometric identification. The Face and Ocular Challenge Series (FOCS) dataset used in the experiment yielded 93.5% Rank-1 accuracy for frontal gait identification and 82.92% Rank-1 accuracy for low-resolution face identification. The score-level multimodal fusion achieved 95.9% Rank-1 identification, demonstrating its superiority and robustness.

Table VI: Proposed Result Analysis with different data sets and Fusion Technique

Dataset	Method	Accuracy	Precision	Recall	F1-Score
ORL	CNN + LSTM	96.8	96.5	96.7	96.6
CASIA-A	CNN + LSTM	91.4	91.0	91.2	91.1
(Deep Fusion)		99.3	99.1	99.2	99.1
CASIA-WebFace	CNN + LSTM	97.6	97.4	97.5	97.4
CASIA-B	CNN + LSTM	93.1	92.8	92.9	93.1
Deep Score Fusion		99.89	99.1	99.0	99.8

Table VI: The proposed model achieves 99.89% accuracy, outperforming all existing methods in Table IV. Traditional machine learning approaches, such as GMM (94.5%) and SVM (98.0%), exhibit comparatively lower performance due to limited feature representation. Deep learning-based approaches improve performance; however, unimodal or partially fused models still lag behind. The superior performance of the proposed HDL-MBF framework is

attributed to the effective integration of spatial (CNN) and temporal (LSTM) features, along with optimised deep score-level fusion. This demonstrates the robustness and efficiency of the proposed approach in multi-biometric identification.

Modalities	Model	Fusion Strategy	Performance / Results	Ref.
EEG signals, affective states(behavioral/physiological)	CNN + LSTM and CNN + GRU	Hybrid (spatial via CNN, temporal via RNN)	Very high identification: up to **99.90-100%** mean Correct Identification Rate (CRR) with CNN-GRU / CNN-LSTM on 40 subjects	[26]
EEG signals for emotion classification	CNN + LSTM (or Bi-LSTM)	Detect emotions/states rather than "who" but a functional template: spatial + temporal fusion.	Accuracies around **97-98%** across several emotion classes on DEAP; ~93.7% on SEED dataset	[27]
Ear image data from multiple datasets	CNN + Bi-LSTM	The hybrid model combining spatial + temporal (bidirectional LSTM) processing of CNN features	Identification rates on datasets: **97.97%**, **99.37%**, **98.57%**, **94.5%**, and **96.87%** depending on dataset	[28]
Signatures (signals + images) + face images	CNN + LSTM hybrid (signal branch + image branch)	Combined features + classification; cross-validation, etc.	Their models (various combinations) perform well; for example, accuracy values in the high 90s (exact numbers depend on the modality pair) — e.g., combining face and speech/signature yields approximately **97.5%** accuracy in some cases.	[29]
Multiple biometric traits (fingerprint, face, voice)	CNN, CNN-GRU, CNN-LSTM compared; CNN-LSTM hybrid	Compared on behavioural, physiological, and voice traits, it also features fusion.	Multifeature fusion identification reached **95.2%**; by individual: face ~84.5%, fingerprint ~82.1%, voice ~67.7%	[30]
Adaptive Fusion Attention (Face and gait)	Deep Neural Networks	Feature fusion	Accuracy 90%(CASIA-A dataset)	[31]
Multimodal (face + gait + body shape) in challenging long-range/degraded imagery	BRIAR dataset; also evaluated in NIST RTE Face in Video Evaluation (FIVE) using BRIAR data; but no mention in that paper of performance on e.g., CASIA-B, Gait3D, etc.	Score-level QME	Accuracy 98.7%	[32]
Gait only	Dynamic Aug Mentation Module (DAM), Temporal Aggregation (TA), Horizontal Mapping (HM)	Global Feature Extractor (GFE) and Dynamic Feature Extractor	GREW; Gait3D; CASIA-B; OU-MVLP (71.4% on GREW, 66.3% on Gait3D, 98.4% on CASIA Band 98.3% on OU-MVLP)	[33]
Multiple biometric traits(face, Gait)	CNN-LSTM hybrid(HDL-MBF)	Deep score fusion	Very high identification: up to **99.89%**	**Proposed work**

Table VII: ACCURACY COMPARISON WITH SOME OTHER EXISTING APPROACHES

Table VII compares CNN- and LSTM-based research in identification. This research achieves high accuracy using a hybrid model comparison of face and gait traits, and Table VI compares the hybrid CNN-LSTM model across different Modalities.

4.3 Limitations

Despite achieving high accuracy or strong performance, the proposed hybrid multi-biometric framework (HDL-MBF) approaches have certain limitations.

First, the dataset may limit the framework (HDL-MBF)'s ability to generalise to large-scale real-world deployments. Second-class imbalance is present across subjects, potentially biasing the classifier. F1 score and confusion matrix calculation reduced this balanced performance evaluation. Third is the issue of images/videos captured in mid or dim lighting, pose changes, and gait-cycle inconsistencies, which directly affect accuracy. This proposed hybrid framework (HDL-MBF) introduces increased computational complexity, longer training times, and higher memory requirements for feature extraction and score fusion. This may restrict deployment in resource-constrained or real-time systems is $O(N^2)$.

5. CONCLUSION AND FUTURE SCOPE

This paper examines the use of Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) to evaluate facial and gait patterns for identity verification using the VGG16 model. The proposed LSTM model uses multiple CNNs to extract spatial features from face frames and each gait frame, and then LSTM networks sequentially capture the gait pattern and temporal changes. By combining the two models, we obtained an identification system that combined physiological and behavioral biometrics, achieving 99.89% accuracy. We evaluated our experiments to see how well it is accomplished alone and with high accuracy when used together with a face and gait model. The F1-score indicate that both models can successfully identify individuals when used together. With continued research and development, this approach will enhance the security of biometric systems.

The dataset is available on the open-source CASIA-B dataset.
http://www.cbsr.ia.ac.cn/english/Gait%20Databases.aspx
Face dataset CASIA-WebFace | Kaggle

REFERENCES

[1] V. Rani and M. Kumar, "Human gait identification: A systematic review," *Multimed Tools Appl*, vol. 82, no. 24, pp. 37003–37037, Oct. 2023, doi: 10.1007/s11042-023-15079-5.

[2] Y. Ojha, "Artificial Intelligence in Armed Conflict: Perspectives from International Humanitarian Law," *SSRN Electronic Journal*, 2025, doi: 10.2139/ssrn.5162209.

[3] R. Liao, Z. Li, S. S. Bhattacharyya, and G. York, "PoseMapGait: A model-based gait identification method with pose estimation maps and graph convolutional networks," *Neurocomputing*, vol. 501, pp. 514–528, Aug. 2022, doi: 10.1016/j.neucom.2022.06.048.

[4] A. Kale, A. K. Roychowdhury, and R. Chellappa, "Fusion of gait and face for human identification," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, pp. V-901–4. doi: 10.1109/ICASSP.2004.1327257.

[5] G. Shakhnarovich, L. Lee, and T. Darrell, "Integrated face and gait identification from multiple views," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Identification. CVPR 2001*, IEEE Comput. Soc, pp. I-439-I–446. doi: 10.1109/CVPR.2001.990508.

[6] C. Shen, S. Yu, J. Wang, G. Q. Huang, and L. Wang, "A Comprehensive Survey on Deep Gait Identification: Algorithms, Datasets, and Challenges," *IEEE Trans Biom Behav Identity Sci*, pp. 1–1, 2024, doi: 10.1109/TBIOM.2024.3486345.

[7] V. Rajasekar *et al.*, "Enhanced multimodal biometric identification approach for smart cities based on an optimized fuzzy genetic algorithm," *Sci Rep*, vol. 12, no. 1, p. 622, Jan. 2022, doi: 10.1038/s41598-021-04652-3.

[8] A. Prakash, S. Thejaswin, A. Nambiar, and A. Bernardino, "Adapt-FuseNet: Context-aware Multimodal Adaptive Fusion of Face and Gait Features using Attention Techniques for Human Identification," in *2023 IEEE International Joint Conference on Biometrics (IJCB)*, IEEE, Sep. 2023, pp. 1–10. doi: 10.1109/IJCB57857.2023.10448765.

[9] C. Chen, Z. Du, L. He, Y. Shi, J. Wang, and W. Dong, "A Novel Gait Pattern Identification Method Based on LSTM-CNN for Lower Limb Exoskeleton," *J Bionic*

Eng, vol. 18, no. 5, pp. 1059–1072, Sep. 2021, doi: 10.1007/s42235-021-00083-y.

[10] N. Senthilkumar, M. Manimegalai, S. Karpakam, S. R. Ashokkumar, and M. Premkumar, "Human action identification based on spatial–temporal relational model and LSTM-CNN framework," *Mater Today Proc*, vol. 57, pp. 2087–2091, 2022, doi: 10.1016/j.matpr.2021.12.004.

[11] P. M, R. Naidu M, and A. P, "Combining Deep Learning Techniques for Enhanced Human Activity Identification: A Hybrid CNN-LSTM Fusion Approach," in *2024 IEEE International Conference on Contemporary Computing and Communications (InC4)*, IEEE, Mar. 2024, pp. 1–7. doi: 10.1109/InC460750.2024.10649335.

[12] A. Kumar, S. Jain, and M. Kumar, "Comparative Study of Multi-Biometrics Authentication Using Machine Learning Algorithms," in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, IEEE, Mar. 2024, pp. 1–5. doi: 10.1109/ICRITO61523.2024.10522125.

[[13] D. M. Ahmed and B. Sh. Mahmood, "Integration of Face and Gait Identification via Transfer Learning: A Multiscale Biometric Identification Approach," *Traitement du Signal*, vol. 40, no. 5, pp. 2179–2190, Oct. 2023, doi: 10.18280/ts.400535.

[14] I. U. Khan, S. Afzal, and J. W. Lee, "Human Activity Identification via Hybrid Deep Learning Based Model," *Sensors*, vol. 22, no. 1, p. 323, Jan. 2022, doi: 10.3390/s22010323.

[15] V. Sharma, M. Gupta, A. Kumar, and D. Mishra, "Video Processing Using Deep Learning Techniques: A Systematic Literature Review," *IEEE Access*, vol. 9, pp. 139489–139507, 2021, doi: 10.1109/ACCESS.2021.3118541.

16] A. Kumar, S. Jain, and M. Kumar, "Deep Learning based Fusion for a Multi-Biometric Identification Using LSTM," in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ACET61898.2024.10730213.

[17 A. Kumar, S. Jain, and M. Kumar, "Face and gait biometrics authentication system based on simplified deep neural networks," *International Journal of Information Technology*, vol. 15, no. 2, pp. 1005–1014, Feb. 2023, doi: 10.1007/s41870-022-01087-5.

[18] B. A. El-Rahiem, M. Amin, A. Sedik, F. E. A. El Samie, and A. M. Iliyasu, "An efficient multi-biometric cancellable biometric scheme based on deep fusion and deep dream," *J Ambient Intell Humaniz Comput*, vol. 13, no. 4, pp. 2177–2189, Apr. 2022, doi: 10.1007/s12652-021-03513-1.

[19] Moon, H., Bey, O., Boubezoul, A., Oukhellou, L., & Mohammed, S. (2025). Real-Time LSTM-Driven Dynamic Gait Mode Detection for Enhanced Control of Actuated-Ankle-Foot Orthosis. *IEEE Transactions on Robotics*.

[20] S. A. F. Manssor, S. Sun, and M. A. M. Elhassan, "Real-Time Human Identification at Night via Integrated Face and Gait Identification Technologies," *Sensors*, vol. 21, no. 13, p. 4323, Jun. 2021, doi: 10.3390/s21134323.

[21] X. Wang and W. Q. Yan, "Human Gait Identification Based on Frame-by-Frame Gait Energy Images and Convolutional Long Short-Term Memory," *Int J Neural Syst*, vol. 30, no. 01, p. 1950027, Jan. 2020, doi: 10.1142/S0129065719500278.

[22]Gait data set The dataset is available on the open-source CASIA-B dataset. http://www.cbsr.ia.ac.cn/english/Gait%20Databases.aspx

[23]Face dataset CASIA-WebFace | Kaggle

[24] Ravina Gupta, Sarika Jain, and Manoj Kumar, "Thermal Vision for Airfield Safety: A Dynamic FOD Detection Framework," *IAENG Int J Comput Sci*, vol. 52, no. 4, pp. 1178–1186, 2025.

[25] J Zhai, Y Xu, and X Yan, "Heterologous Image Matching Based on Salience Region," *IAENG Int J Comput Sci*, 2025.

[26] W. Meng *et al.*, "Emotion identification via affective EEG signals: State of the art," *Neurocomputing*, vol. 643, p. 130418, Aug. 2025, doi: 10.1016/j.neucom.2025.130418.

[27] A. Mahajan and S. K. Singla, "DeepBio: A Deep CNN and Bi-LSTM Learning for Person Identification Using Ear Biometrics," *Computer Modeling in Engineering & Sciences*, vol. 141, no. 2, pp. 1623–1649, 2024, doi: 10.32604/cmes.2024.054468.

[28] S. Salturk and N. Kahraman, "Deep learning-powered multimodal biometric authentication: integrating dynamic signatures and facial data for enhanced online security," *Neural Comput Appl*, vol. 36, no. 19, pp. 11311–11322, Jul. 2024, doi: 10.1007/s00521-024-09690-2.

[29] M. Ramzan and S. Dawn, "Fused CNN-LSTM deep learning emotion identification model using electroencephalography signals," *International Journal of Neuroscience*, vol. 133, no. 6, pp. 587–597, Jun. 2023, doi: 10.1080/00207454.2021.1941947.

[30] Z. W. Wenrui Shen, "Person Re-Identification Algorithm Based on Improved ResNet," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 5, 2024.

[31] T. S, A. Prakash, A. Nambiar, and A. Bernadino, "Exploring Fusion Techniques and Explainable AI on Adapt-FuseNet: Context-Adaptive Fusion of Face and Gait for Person Identification," *IEEE Trans Biom Behav Identity Sci*, vol. 6, no. 4, pp. 515–527, Oct. 2024, doi: 10.1109/TBIOM.2024.3405081.

[32] F. et al. Liu, "Person Identification at Altitude and Range: Fusion of Face, Body Shape and Gait," 2025.

[33] M. et al. Wang, "DyGait: Exploiting Dynamic Representations for High-performance Gait Identification," *Proceedings of the IEEE/CVF international conference on computer vision. 2023.*, 2023.

[34] S. Naoui, M. H. Elhdhili and L. A. Saidane, "Novel Smart Home Authentication Protocol LRP-SHAP," *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, 2019, pp. 1-6, doi: 10.1109/WCNC.2019.8885493

Author's biography

1. Amit Kumar – Amit Kumar is a research scholar at Amity University, Noida, India
2. Dr Sarika Jain - Professor at Amity University, Noida, India.
3. Dr Manoj Kumar -Professor at School of Computer Science, University of Wollongong, UAE